

-52-

5 We Claim:

1. A method for inferring a gene network, comprising
 - (a) providing an inferential model of possible gene networks of an organism including defining a search space;
 - 10 (b) selecting a biologically relevant subspace of said search space; and
 - (c) calculating an optimal solution in said selected subspace by repeatedly applying an algorithm that computes small gene networks optimally.
2. The method of claim 1, wherein said inferential model is a Bayesian network

15 estimation model.
3. The method of claim 1, wherein said biologically relevant subspace includes genes relating to a metabolic pathway of said organism.
- 20 4. The method of claim 1, wherein said algorithm comprises the steps:
 - (a) compute $F(g, \phi) = s(g, \phi)$ for all $g \in G$;
 - (b) for all $A \subseteq G$, $A \neq \phi$ and all $g \in G$ compute $F(g, A)$ as $\min\{s(g, A), \min_{a \in A} F(g, A - \{a\})\}$;
 - (c) set $M(\phi) = \phi$;
 - 25 (d) for all $A \subseteq G$, $A \neq \phi$, do the following steps:
 - (i) compute $g^* = \arg \min_{g \in A} (F(g, A - \{g\}) + Q^{A - \{g\}}(M(A - \{g\})))$; and
 - (ii) for all $1 \leq i < |A|$, set $M(A)(i) = M(A - \{g^*\})(i)$, and $M(A)(|A|) = g^*$; and
 - (e) return $Q^G(M(G))$.
- 30 5. The method of claim 4, wherein said algorithm is modified according to the steps of:
 - (a) in the computation of F in Step 1 and Step 2, compute only $F(g, A)$ for all $g \in S_i$ and all $A \subseteq C_g$; and

-53-

- 5 (b). replace the term $F(g, A - \{g\})$ in Step 4a by $F(g, (C_g = S_i) \cup (C_g \cap A))$.
6. The method of claim 1, wherein an optimal network N has a definition: $score(N) = \sum_{g \in G} s(g, P^N(g))$.
- 10 7. The method of claim 1, wherein said algorithm comprises the steps:
- (a) cluster genes in G such that no cluster is larger than c genes;
 - (b) sort the clusters by decreasing size: C_1, \dots, C_n ;
 - (c) for each $i \in \{1, \dots, n\}$ and for each $g \in C_i$, select up to m candidate parents from $C_1 \cup \dots \cup C_n$; and
 - 15 (d) compute an optimal gene network model using Theorem 1.2.
8. The method of claim 1, wherein said algorithm comprises the steps:
- (a) group genes in G in groups C_i with $|C_i| \leq c$ and sort them according to biological knowledge: C_1, \dots, C_n ;
 - 20 (b) for each $i \in \{1, \dots, n\}$ and for each gene $g \in C_i$, select up to m candidate parents from $C_1 \cup \dots \cup C_i$; and
 - (c) compute an optimal gene network model using Theorem 2.
9. The method of claim 1, wherein said algorithm comprises the steps:
- 25 (a) compute $F(g, \phi) = s(g, \phi)$ for all $g \in G$;
 - (b) for all $A \subseteq G$, $A \neq \phi$ and all $g \in G$ compute $F(g, A)$ as $\min\{s(g, A), \min_{a \in A} F(g, A - \{a\})\}$;
 - (cc) set $M(\phi) = \phi$;
 - (d) for all $A \subseteq G$, $A \neq \phi$, do the following two steps:
 - 30 (i) compute $g^* = \arg \min_{g \in A} (F(g, A - \{g\}) + Q^{A - \{g\}}(M(A - \{g\})))$; and
 - (ii) For all $1 \leq i < |A|$, set $M(A)(i) = M(A - \{g^*\})(i)$, and $M(A)(|A|) = g^*$; and
 - (e) return $Q^G(M(G))$.
10. The method of claim 1, wherein said algorithm comprises the steps:

-54-

- 5 (a) set $F^m(g, \phi, 1) = \phi$, $S^m(g, \phi, 1) = s(g, \phi)$ for all $g \in G$;
 (b) for all $g \in G$, all $A \subseteq G$, $A \neq N$ and all $n \leq m$ do the following two steps:
 (i) select $B^* \subseteq A$ from $\{B \subseteq A \mid B = A \vee B = F^m(g, A - \{h\}, p), h \in A, p \leq m\} - \{F^m(g, A, p) \mid p < n\}$ such that $s(g, B^*)$ is minimized; and
 (ii) set $F^m(g, A, n) = B^*$, $S^m(g, A, n) = s(g, B^*)$;
 10 (c) set $M^m(\phi, 1) = \phi$ and $D^m(\phi, 1) = \phi$;
 (d) for all $A \subseteq G$, ϕ , and all $n \leq m$ do the following three steps:
 (i) choose a triple $(g, p, q) \in A \times IN_{sm} \times IN_{sm}$ such that $score(Q^{A-\{g\}}(M^m(A - \{g\}, p), D^m(A - \{g\}, p))) + S^m(g, A - \{g\}, q)$ is minimized and (g, p, q) induces a network different from $Q^A(M^m(A, r), D^m(A, r))$ for $r < n$;
 15 (ii) set $M^m(A, n)(i) = M^m(A - \{g\}, p)(i)$ for $i < |A|$, and $M^m(A, n)(|A|) = g$;
 and
 (iii) let v denote $D^m(A - \{g\}, p)$. Set $w \in IN^{|A|}$ as $w_i = v_i$ for all $i < |A|$ and $w_{|A|} = q$ and set $D^m(A, n) = w$; and
 (e) return $Q^G(M^m(G, i), D^m(G, i))$ for all $i \leq m$.
 20
11. The method of any of claims 1-10, wherein reliability of an enumerated gene network, comprising the steps:
- (a) enumerate the most likely gene network models N_i , $1 \leq i \leq n$;
 (b) for every $g, h \in G$, count the occurrences of the edge (g, h) in the networks
 25 N_i ;
 (c) select all edges (g, h) with at least c occurrences;
 (d) for all subsets M of the set of selected edges with $|M| = k$, count the networks including all edges in M ; and
 (e) return all motives M with at least c occurrences.
 30
12. The method of any of claims 1-11, further comprising calculating a scoring function selected from the group consisting of BRNC score, BDe score and MDL score.
- 35 13. A method for determining a gene network as substantially described herein.
14. A storage medium containing results obtained using the method of any of claims 1-11.

-55-

5 15. A storage medium containing results obtained using a method as substantially described herein.

16. A system for determining gene network relationships, comprising:

an input device for providing quantitative expression data for genes of an organism;

10 a storage device adapted to receive quantitative expression data for genes of said organism;

a processor adapted to carryout a Bayesian network analysis of network relationships between said genes, thereby producing a data set reflecting said network relationships; and

an output device for displaying said data set of said network relationships.

15

17. A system for determining gene network relationships as substantially described herein.